

# OPTIMIZING SERVER DELIVERY OF CONTENT BY SELECTIVE INCLUSION OF OPTIONAL DATA BASED ON OPTIMIZATION CRITERIA

## INVENTOR

Taylor S. Gautier

## BACKGROUND

### Field of Invention

The present invention relates generally to systems, methods, and software products for delivering content in client-server networks, and more particularly, to systems, methods, and software products that optimize the delivery of such content in response to optimization criteria.

### Background of the Invention

The Internet generally, and the World Wide Web specifically, provide a client-server model of content delivery, in which content stored on a server at a web site is delivered over a communications channel to a client device. Conventionally, the content as structured by the author is delivered by the server to all requesting clients in the same form. That is, the server does not change the structure of the content to accommodate the performance characteristics of the communication channel to a particular client, or the performance characteristics of the client itself. At best, the client re-formats the received content based on its own capabilities (e.g. a text only browser will filter out images, but still must receive the images, thereby consuming bandwidth).

This delivery model is not satisfactory when there are many different types of clients, with varying performance capabilities and varying communication channels. For example, today a web site may be accessed by many different types of clients, from conventional full size computers with large screen displays (e.g. 13" to 17") and full color capabilities (16M colors) to small handheld computers with small screens (e.g. 4" to 6") and limited color capabilities (256 colors), to cellular telephones with very small screens (1" -2") and virtually no color capabilities (2 - 8 colors monochrome). A web page with many images that is designed to be viewed on a full size, full color monitor will not appear at all properly on a cellular telephone display with only monochrome

output. Moreover, the cellular telephone has a much lower bandwidth capability than a standard desktop computer, and thus would require significantly more time (and hence more cost and delay) in loading such a web page.

Typically, in order to accommodate these varying capabilities, a server must  
5 maintain many different web sites or web pages, each designed to be served to a particular class of client. But this approach is very labor intensive, particularly for sites that must be updated frequently, since the many different sizes and content selections for a given page must be kept in synchrony to all have the same current content. Further, display capabilities are only one characteristic that affects client performance.  
10 Different client devices may operate with different bandwidth capabilities, having different processor and memory capacities, or even have different pricing schemes associated with retrieval of content. When considering the possible combinations of factors that affect client performance, such as bandwidth and display capabilities (to name just two), the number of different possible web sites that need to be maintained  
15 quickly becomes unmanageable.

An alternative solution to maintaining multiple web sites is to design web sites that are acceptable to only the lowest performance devices that are going to access the site. But this approach yields simple sites that do not fully allow higher performance devices to take advantage of their performance features.

Accordingly, it is desirable to provide a system, method, and software product  
20 that can optimize delivery of content to client devices in response to optimization criteria. In particular, it is desirable to provide a web server that can efficiently deliver content to client devices having varying bandwidth, display, and other performance characteristics.

25

## SUMMARY OF THE INVENTION

The present invention overcomes the limitations of the conventional systems by providing a server software product and a method that selectively delivers content to a client in response to optimization constraints indicative of the client's performance characteristics. The server delivers content such as web pages, images, text, audio,  
30 video, and any other data types to varying client devices with varying performance characteristics.

The present invention augments existing markup languages with a new tag that demarcates a group of one or more content items that may be optionally included in the content to be delivered. Where there are multiple optional items in a group, these items are alternatives for inclusion in the content, and one of the items is selected by the server for inclusion in the content based on an optimization constraint dependent on the characteristics of the client device.

Preferably, the items are ordered in a manner corresponding to some set of client characteristics; one of these items is selected as the content to be delivered when the content is requested. The characteristics may be performance characteristics of the client or communication channel, or other attributes of the client. For example, the ordering of optional content items may be with respect to highest performance to lowest performance devices, or vice versa. For example, in web page providing directions to a desired location, a group of optional content may include first an image of a map illustrating the directions, followed by a detailed text description of the directions, followed by a very short text description of the directions. Thus, each of these items is intended for different types of client devices, which can best receive and display the directions in one of these formats, and one of these forms of information will be delivered to a requesting client.

When the server is processing a client request for the document, any content that is demarcated as being optional is included in the delivered content depending on the performance characteristics of client that is requesting the content. More particularly, the server processes the items of optional content in the order in which they are demarcated together. Thus, in an embodiment in which optional items are ordered from highest to lowest performance capability, if the requesting client is a high performance device (including having a high bandwidth connection), then the first item of content is included in the document; if the requesting client is a low performance device, then the first item(s) is skipped until the appropriate item of content is reached. Obviously, the reverse of this ordering and selection process is possible. The selection of which item of content is appropriate for which client device is based on an optimization constraint. The server stores data which associates different types of client devices or performance characteristics with different levels of the optimization

constraint. The number of different types of client devices or performance characteristics is not limited. The server selects the appropriate optimization constraint based on observable characteristics of the client device requesting the content. These observable characteristics include the device type, browser type, operating system,  
5 processor, memory, user preferences, software installed on the client, and language type for the requested content.

A set of selection rules is used by the server to select which of the optional items of content are to be delivered to the client device based on the optimization constraint. More particularly, the rules map any optimization constraint to one of the items of  
10 content. In a preferred embodiment, where the optional items are ordered with respect to their performance demands, the optimization constraint is used as an index into these items to select the appropriate one for the client. The process may be understood as one of selectively reducing the data stream provided to a specific client based on encoding of optionally selectable/filterable data, and an optimization constraint value  
15 that controls the selection process.

The tag that demarcates optional content can obviously be any useful string. In one embodiment, the tag is usefully called the "OPT" tag. The OPT tag is formatted using standard SGML formats, with <OPT> indicating the beginning of the optional content, and terminated by </OPT> indicating the end of the optional content. Within  
20 the OPT tag there may be zero or more delimiting tags separating alternative items of optional content. The preferred delimiting tag for separating alternative items is <DOPT> for "Delimiting OPTion". A well formed OPT tag has the following form:

<OPT> *content* {<DOPT>*alternative content*}\*</OPT>

Zero or more <DOPT> tags with their corresponding *alternative content* follow in  
25 an order, which may be from highest to lowest performance requirements, or vice versa. In the former ordering, this typically means from largest data size (e.g. length of text or use of images) to smallest data size.

Examples:

1. <OPT>Real Time Stock </OPT>Quotes
2. <OPT>Turn left<DOPT>Left<DOPT>L</OPT> on First Street

In the first example, the optional content "Real Time Stock" is delivered only to client devices that have no performance limitations, e.g. are not bandwidth limited, and so receive the complete text "Real Time Stock Quotes". A client device that has any optimization constraint only receives "Quotes" when this content is delivered. Thus, the unconstrained version is appropriate, for example, for a standard desktop computer, but the latter version is better supplied to a device with a very small screen or low bandwidth capabilities, such as a cellular telephone.

In the second example, of directions being given to a location, the alternative content selections become progressively smaller, being ordered left to right from least constrained (highest performance) to most constrained (lowest performance). Thus, the first item "Turn left" is delivered to a high performance client, while the last item "L" is suitable for client devices with lower bandwidth and/or smaller displays.

One preferred method of using the optimization constraint as an index into the optional items of content uses both positive and negative numbers as optimization constraints. Positive numbers index from least constrained items to most constrained items. In example 2) above, an optimization constraint of (1) would select "Turn left" while an optimization constraint of (2) would select "Left". Negative numbers index from most constrained to least; thus (-1) would select "L", which (-2) would also select "Left". Preferably, positive one (1) always selects the least constrained item, and negative one (-1) selects the most constrained item; zero (0) is treated as negative one (-1). It is preferable rule that a negative number may never select the least constrained most item, and a positive number may never select the most constrained item. With these rules, any number of optional items of content may encoded in the document and selectively delivered. Again, whether the ordering is left to right or right to left for high to low constraint is an implementation detail; which ever is used, the selection logic is implemented to match.

Using the present invention, it is possible to build a single server which is capable of serving information to multiple clients, over varying bandwidth connections, or with other varying performance characteristics.

The amount of effort required to support a growing number of clients is reduced from an order of  $O(2)$  [for a traditional system], or a square relationship, to an order of  $O(1)$ , or a linear relationship.

This savings of effort required to maintain such a server brings the possibility of  
5 serving data to a wide range of client devices over a broad scale of bandwidth connections to a reality.

09312506-05149

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is an illustration of a system in accordance with the present invention.

Fig. 2 is an illustration of a method of operating a server in accordance with the present invention.

5 Figs. 3a-3c illustrates an example of an encoded content item with optional content, and different outputs of optional content according to varying levels of optimization constraint.

16331/03528/781864 V

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Referring to Fig. 1, there is shown an illustration of a system in accordance with the present invention. The system includes a server 102 and a content database 100. The server 102 of interest is the software product that executes on a conventional server  
5 grade computer, such as a workstation, minicomputer, mainframe, or the like. The server 102 is configured to execute in accordance with the functional features as described herein.

The content database 100 stores content in a suitable format, such as a markup language, ASCII, and the like. Suitable markup languages include HyperText Markup  
10 Language (HTML), Extended Markup Language (XML), Hypertext Dynamic Markup Language (HDML), Tagged Text Markup Language (TTML). The server 102 is then a server of the appropriate type of markup language, and includes functions for receiving requests for a content item, typically indicated by a uniform resource locator (URL), and parsing markup language tags, creating a content item in response to a request, and  
15 transmitting the content item to the requesting client.

Some of the content in the database 100 is structured to include at least one item of optional content demarcated by a specific markup language tag. The preferred tag is <OPT>, as described above, and its terminating tag </OPT>. Any number of alternate items of content may be enclosed, each delimited by a specific tag, such as <DOPT>.  
20 The items are preferably ordered with respect to some performance criteria, such as bandwidth capability. One ordering is from left to right, from lowest constraint, highest performance devices, to highest constraint, lowest performance devices. Schematically, a portion of content structured in this manner is as follows:

<OPT>*lowest constraint data*<DOPT>*next lowest constraint*<DOPT>....<DOPT>*highest*  
25 *constraint data*</OPT>.

In one embodiment, these tags are part of meta-language that itself marks up other markup languages. In this embodiment, the OPT and DOPT are set off by << and >>, to indicate to the server 102 that they are part of the meta-language, and not part of the underlying markup language. In this manner, any number of different markup  
30 language content items are stored in the database 100, and they are marked up in the meta-language. Various different client devices use various different ones of the



markup languages. Client device specific pages are thus constructed in the different markup languages. These client specific pages are marked up with the meta-language tags, and here particularly the OPT tag is used to specify optional content that can be merged or integrated into the client specific pages. In a preferred embodiment, the content within the OPT tags is variable data selected from the database 100 in response to a query; the server filters the query response using the OPT mechanism described herein. The server 102 can thus deliver content in any of the markup languages, and with any of the optionally defined content to any requesting client; this allows different clients which use different markup languages to interface with the server 102.

The server 102 communicates with various types of client devices 106 over a communications network 104. The client devices 106 include desktop and laptop computers, pocket organizers, personal digital assistants, cellular telephones, and any other type of communication device adapted for Internet or Web based communications (i.e. HTTP). The server 102 selectively provides optional content from a content item, depending on an optimization constraint associated with a requesting client device 106.

Referring to Fig. 2, there is shown a method in accordance with the present invention. The server 102 receives 200 a request for a content item from the database 100. The server 102 determines 202 an optimization constraint associated with the requesting client device. The server 102 compiles 204 the requested content item, including selectively including content that is set forth as optional content, and if necessary, by selecting one of the alternate items of content, based on the optimization constraint. The server 102 then transmits 206 the compiled content item to the client. Determining the optimization constraint preferably includes determining either a client device type, client operating system type, client browser type, content language type of the requested content, communication bandwidth of the client, client processor, memory, user preferences, or software installed on the client, or a combination of any of these or other attributes. Selecting one of the alternate items preferably includes applying a set of rules that map the optimization constraint to selected ones of the optional content items.

One set of rules uses the optimization constraint as an index, where the items are ordered in manner corresponding to performance capabilities of the client devices. This set of rules is described as follows:

1. Determine the number of optional items. If only one is available, add a blank option as the rightmost (most constrained) option.
2. Retrieve the optimization constraint which corresponds to the client device.
3. Positive numbers select from least constrained to most constrained. Negative numbers select from most constrained to least constrained.
4. The positive number one (1) selects the least constrained item.
5. The negative number one (-1) selects the most constrained item.
6. The number zero (0) is equivalent to negative one (-1)
7. Use the optimization constraint as an index into the optional items, where 1 selects the least constrained item, 2 the second least constrained item, and so on; -1 selects the most constrained item, -2 selects the second to most constrained item and so on.
8. A negative number may never select the least constrained item.
9. A positive number may never select the most constrained item.

The following table illustrates an example mapping where there are 5 items of alternate content, using a left to right, least to most constraint ordering:

Table 1

	1 <sup>st</sup> Item (Least Constrained)	2 <sup>nd</sup> Item	3 <sup>rd</sup> Item	4 <sup>th</sup> Item	5 <sup>th</sup> Item (most constrained)
Opt. Constraint $n$ which selects this item:	$n=1$	$n=2,$ $n \leq -4$	$n=(3, -3)$	$n \geq 4,$ $n=-2$	$n=(-1, 0)$

In one embodiment, the optimization constraint is defined with respect to the markup language of the content being served. The next table defines one set of optimization constraints for this embodiment:

Table 2

Language/Bandwith	Optimization Constraint
HTML (on T1 connection at 1.5Mb/s)	1
HTML (wired modem <=56Kb/s)	2
HTML (wireless <=9600 bps)	-3
HDML (<9600 bps)	-2
TTML	-1

The assignment of optimization constraints based on content language is a reflection of the typical bandwidth limitations of client devices that use the specific content language. HTML is generally used in a desktop environment where connection speeds are relatively high and display capacities are also relatively high, and thus a low constraint is appropriate, particularly for HTML on a T1 connection. HTML on a wireless device, such as a PDA is more constrained, and hence a higher constraint. HDML is generally used in a digital cellular telephone communications arena, with a transmission speed limited to 9600 bps and size of messages limited to the telephone memory, hence even more constrained environment, as reflected by the optimization constraint. TTML is generally used over GSM style SMS messages which are limited to 160 characters, and thus is the most constrained environment, and thus given a -1 optimization constraint.

As the server 102 scans an item of content, it comes across an OPT encoded item of data. The server 102 can determine the optimization constraint based on the markup language as in Table 2, or using any other set of rules mapping information about the client device attributes to an optimization value. The determined value is then used as an index into the items of DOPT portions, to select the appropriate item of content. This process is repeated for each OPT encoded portion of the requested content, so that the server 102 ultimately compiles a complete content item. The completed content item is transmitted to the client device.

Figs. 3a-3c illustrates a more complex example of the present invention. Here, a content item 300 in Fig. 3a is encoded in a markup language, along with multiple portions set out as optional content between OPT tags, some of which have multiple alternate items delimited by DOPT tags. Fig. 3b illustrates the output of this content item by selection for a least constrained environment, HTML over T1. Note that all of

